NPS-55Bn-La9091A

UNITED STATES
NAVAL POSTGRADUATE SCHOOL

OBJECTIVE IDENTIFICATION PROCEDURES FOR THE NAVAL OIL ANALYSIS PROGRAM

by

Donald R. Barr

and

Harold J. Larson

September 1969

This document has been approved for public release and sale;
its distribution is unlimited.

UNITED STATES NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Robert W. McNitt, USN
Superintendent

Dr. Robert F. Rinehart
Academic Dean

ABSTRACT:

A class of rules is developed for making decisions concerning whether a mechanical system may be failing, based upon spectroscopic analyses of the system's oil over a period of time. Some considerations that went into the development of these rules, including conclusions based upon studies of certain analysis records and experiments, are presented. It is indicated that these identification procedures should perform well in connection with a computerized analysis system, at least insofar as routinely monitoring the "well behaved" systems, while calling the attention of appropriate personnel to possibly discrepant systems.

TABLE OF CONTENTS

OBJECTIVE IDENTIFICATION PROCEDURES FOR THE NAVAL OIL ANALYSIS PROGRAM

H. J. Larson and D. R. Barr
Naval Postgraduate School

I. INTRODUCTION

The Navy Oil Analysis Program (NOAP) was begun in 1956 as an investigation of the practicality of the use of spectrometric analysis of the circulating oil from aircraft engines in describing the internal condition of the engines. The initial program was small with relatively few aircraft involved; by 1958 the program had proved beneficial and the effort was considerably expanded. At present the intention seems to be to involve virtually all Navy fluid lubricated mechanical systems in the program.

Since this report is mainly concerned with investigations for Navy aircraft, the following working descriptions will be limited to procedures used for aircraft engines. Currently, reciprocating engines participating in NOAP are sampled roughly every 30 hours and participating jet engines are sampled roughly every 10 hours; the sampling is accomplished after the aircraft has returned from a flight and before the oil has gotten cold. A special sampling kit is provided for each specific engine to be sampled. This kit generally consists of a sampling tube and a sample bottle; the sampling tube has been cut to a predetermined

length so that, if it is inserted into the oil reservoir in a prescribed manner, it will not pick up sludge from the bottom of the reservoir. After the tube has been inserted into the oil long enough for oil to enter the tube, the top end is stopped with the operator's finger and the contents transferred to the sampling bottle. The sampling bottle is then capped and mailed to the laboratory, together with a sheet listing the unit model number, the unit serial number, date of the sample, hours since oil change and hours since overhaul of the engine.

When the sample is received at the laboratory it is carefully recorded and, depending on the number of arriving samples, analyzed almost immediately on the spectrometer. The spectrometer has two carbon electodes, one a stationary pencil and the other a rotating disk. When a sample is to be analyzed on the spectrometer the cap of the sampling bottle is almost filled with the sample oil. Then the rotating disk is placed in the oil in the cap, the gap between the two electrodes is set, the disk electrode is started rotating at 30 rpm and an arc is fired across the gap for roughly 25 seconds, burning the oil carried to the uppermost side of the rotating disk. The light from the burning oil is analyzed simultaneously for the intensity of the characteristic spectral lines of 10 elements, commonly those of aluminum, copper, iron, magnesium, nickel, silver, chromium, tin, silicon and titanium. By referencing these intensities to a built-in standard, the spectrometer translates these "average" intensities into readings in parts per million for the various elements. These readings are

then automatically recorded on a punched card containing previously hand-entered information identifying the sample and the date it was analyzed.

The sampling kit materials are all discarded after one use, as is the rotating disk electrode, to avoid contamination of one sample by another. Also, the pencil electrode is reshaped in a sharpener after each use, to prevent any splashed oil from affecting the readings for a subsequent sample. Generally, only a portion of the oil received in the sampling bottle is consumed in the analysis and the remaining oil is discarded.

The sample readings in ppm of the various elements are used as an aid in deciding what the internal condition of the engine may be. Presumably, if the engine is in good operating condition, the true amount of contamination in the circulating oil should be within prescribed "normal" limits at any given time and the amount of contaminants added to the oil between sampling periods should also lie within "normal" limits. Thus, if the indicated level of contaminants and the rate of increase of contaminants are in the normal range, no action is taken and sampling continues at the normal rate. If, however, either the indicated level of one or more contaminants or the rate of increase of sample readings of one or more contaminants (since the last previous sample from the same engine) lie above the normal values, some action will be taken by the lab. Generally, a check sample is gotten first, to verify the high readings, and then, if the high readings are verified, either the aircraft is grounded and maintenance is requested or it is

requested that future samples be taken more frequently (for example, sample every 10 hours rather than every 30 hours). Which of these two actions is taken is subjective and is related to how high the level of contaminant or the rate of increase is above normal. The "normal levels" for each model are evolved subjectively over time both from engineering test data, supplied by the engine manufacturers prior to a new model being placed in service, and from accumulated operational data with the particular model after it has been placed in general use. See [3] for a more detailed description of the history of NOAP and of current procedures.

The present report is concerned with an explanation of a statistical analysis which might be used on the spectrometer readings to objectively identify those aircraft requiring special action. Succeeding sections will discuss the inherent errors of the sampling procedure and of the spectrometer readings, the results of some preliminary analyses on spectrometric oil analyses furnished by the Navy lab at Pensacola. These are used in turn to formulate and describe a particular analytic technique that could be used for the objective analyses on a working basis.

II. STOCHASTIC NATURE OF THE OBSERVATIONS

1. Introduction

In this section a discussion is given of the inherent variability that is observed if the same oil sample is run on the same spectrometer repeatedly; each reading in such a set of readings of contaminant concentrations is referred to as a trial of an

experiment. Other sources of variability in the observed ppm (parts per million) readings are also discussed and a general model is proposed which might be used to estimate the ppm content of the oil in an engine and deduce the quality of this estimate, based on an observed spectrometric analysis of a sample of the engine oil.

In most situations involving repeated trials of an experiment, the results of the various trials are not precisely the same, but vary from trial to trial. This is usually the case, even though considerable effort is expended in attempting to make the experimental conditions the same for each trial. The experimenter's inability to exactly reproduce a result observed on a previous trial of the experiment, especially when working close to the possible limits of measurement, as in the case of oil analysis, is certainly to be expected. This inherent variability is always observed when measurements are made in extremely fine units.

The amount by which an observed result differs from the "true" theoretical value is called error. One objective usually considered in formulating a theory (or model) to "explain" a phenomenon under investigation is to reduce the error to a tolerable level. For example, an experimenter might be quite willing to take into account only those conditions which affect the outcome in a relatively major way, choosing to ignore the minor ones and clumping their combined effect into error. More commonly, it is impossible to account for all of the factors having an influence on the observations obtained in repeated trials of the experiment. Thus, from a practical point of view, in order to formulate models for most phenomena, we are

forced to use rather naive models which take into account only a few of the great number of factors influencing the outcome of the experiment. This in turn may make the unexplained portions of the values observed (that is, the errors) rather large. Such appears to be the case with spectrometric analysis of used engine oil.

In order to estimate the "true" ppm content of the oil sample when the experimental results include errors, and to estimate how great the errors might be, the results of many experimental trials may be statistically analyzed. Such an analysis usually involves the formulation of a statistical model, which in turn depends on making certain assumptions concerning the random behavior of the errors that might be encountered in repeatedly performing the experiment, together with certain measures calculated from the actual observed results. Two such measures are the sample mean and variance which are estimates of the theoretical expected value of the experimental result and the error (measured from this expected value), respectively.

Before discussing a statistical model for the spectrometric analysis of used engine oil, we pause to discuss some possible sources of error in such analyses. In the present case, the term error, for a certain element, refers to the difference between a value posted in the record file of a listed engine for a certain listed time since overhaul and oil change and the true mean concentration of that element in the oil reservoir of that engine at that time. Of course, since the latter value cannot be observed, we cannot actually measure errors, but rather must make inferences

about their magnitude from statistical analyses of the records of past oil analyses. The following discussion of sources of errors in oil analysis data is not exhaustive, but it is felt that the major sources are included. The errors discussed are grouped into three main categories: non-representativeness of the oil burned, the analysis, and the record-keeping procedure.

2. Errors in the Spectrometric Analyses of Oil

A. Non-representativeness of the oil burned. Since an attempt is being made to make inferences concerning the possible failure of a mechanical system, using the characteristics of the system's oil, it is important that the oil actually analyzed be representative of the oil in the system. Failure to achieve exact representativeness gives rise to error. Let us now discuss a few specific sources of such error.

First, only a small sample of the oil in the reservoir of an engine is actually sent for analysis. Such a small sample might not be exactly representative of the oil in the reservoir for several reasons: the oil in the reservoir may not be homogeneous (one might find, for example, tendency for a slightly higher concentration of iron near the bottom of the reservoir than near the surface of the oil). It is also possible that the process of taking the sample tends to influence its composition, for example through lack of cleanliness in the sampling tube or bottle, or slightly different technique of taking samples by the various people involved. Second, the oil actually burned in the analysis is but a small portion of a sample (poured into the sample bottle

cap) taken from the sample bottle. The overall effect is thus that an extremely small volume of oil is actually burned; this portion hopefully is representative of all of the oil in the reservoir at the time of initial sampling. In addition, there is a chance of contamination of this sample each time the sample oil (and certain parts of the spectrometer "burning" apparatus, discussed below) is handled, up to and including the actual time of burning.

Of course, as outlined in Section I, portions of the sampling and analysis procedure have been designed specifically to reduce these errors as much as possible. There does not seem to be a reasonable way to determine the extent of error remaining, (in spite of procedural steps taken to eliminate them) due only to these possible sources of errors, short of carrying out a carefully planned experiment with this aim in mind.

B. The Analysis. Several potential sources of error can be identified in the analysis procedure and mechanism. These errors can be thought of as giving rise to different analysis results, even if we imagine that the oil poured into the sample homogeneous and truly representative of that in the reservoir from which the sample was taken. Let us consider, then, an analysis of a sample, followed by a second analysis of the same sample at some later time. Some possible causes for getting different results on these analyses, even when it is assumed that the spectrometer is "recalibrated" with a standard before each of the analyses, are as follows.

First, the "strength" of the spectral lines monitored depends

in part upon the volume of oil actually burned in the analysis run. It is impossible to guarantee that this volume is the same on each of the analyses in question (or the same as that in the corresponding calibration runs). This may be due in part to differences in the physical characteristics of the rotating disc, the depth of this disc in the oil in the cap, the speed of rotation of the disc, the viscosity of the oil (which is affected, for example, by the temperature and chemical composition of the oil sample placed in the bottle cap), and the duration of the burn. All of these may change slightly from one analysis to the next.

Second, the strength of the spectral lines may be affected by the size of the gap between the electrodes, and their composition and other physical characteristics (such as shape). Third, the emission of energy by the burned oil is inherently a random phenomenon--the number of atoms of a certain element actually excited, which subsequently emit radiation which arrives at the exit slit in the spectrometer, will theoretically vary from one analysis to another even if the samples and burning conditions are identical and identical amounts of oil are burned in each analysis.

The measurement of the strength of a given spectral line by the signal produced by a photomultiplier tube and the subsequent conversion to a reading in digital form undoubtedly involves some error. Finally, the calibration of the spectrometer according to certain "standard" samples involves error, both because exact standard samples are difficult (if not impossible) to prepare and

maintain, and because the procedure of adjusting the machine to produce output in agreement with the supposed standard concentrations may involve slight errors. In addition, even if the spectrometer and output mechanism were properly calibrated at a given time, this may not be the case at a later time due to changes in the many factors influencing the spectrometer, such as temperature, barometric pressure and humidity.

As in the case of errors due to non-representativeness of the oil burned (Case A), steps have been taken to reduce the overall error due to the analysis, as discussed in Section I. Unlike Case A, however, it is possible to make inferences about the combined effects of errors in analysis. One method of doing this is to observe the results of several analyses of the same sample, perhaps with a standard sample. The data from such an experiment are available (Air Force data), and are discussed in Section III below.

C. The Record-Keeping Procedure. The current method of keeping records of the results of the analyses of oil samples from each specific unit being monitored involves several possible sources of error. For example, the information accompanying a sample sent for analysis includes several entries in a standard form, made "by hand" by someone in the group initiating the sample. These hand entries include the model number and serial number of the engine from which the sample was taken, and the accumulated hours since the engine was overhauled and since the oil was changed in the engine (the latter being presumably taken from records which

are themselves subject to error). For various reasons, then, it is possible that incorrect information may be entered upon the form accompanying the sample. In addition, some of this information is read and punched (typed) into certain data cards maintained at the analysis center. These data cards for the engines identified by the hand entries are "pulled" from a file by personnel at the analysis center. Of course, the combined operations of pulling a data card from a file and entering the handwritten information on it may give rise to error.

There seems to be no realistic way to estimate the magnitude of errors due to the record-keeping procedure without performing an experiment specifically designed for this purpose.

3. A Statistical Model for Repeated Spectrometric Analyses

We shall now discuss a statistical model which appears to provide a reasonable explanation of the apparent errors observed in past spectrometric analyses of used engine oil. In view of the steps taken in the sampling and record-keeping procedures to reduce as much as possible the errors due to non-representativeness of the oil burned and the record-keeping procedure, it seems reasonable that the major portion of the overall errors in the oil analysis program are due to the analysis procedure itself. In what follows, we shall find it convenient to view all errors as arising in the analysis of the oil (Case B).

Suppose, then, that the oil in the engine reservoir is quite homogeneous and that a representative sample of oil has been selected and placed in the sampling bottle. To simplify the

discussion at this time, also assume that only one element, for example iron, is of interest, and that the true iron content of the engine reservoir and of the oil in the sampling bottle is μ ppm. The quantity of oil in the sampling bottle is sufficient to run at least 20 different analyses on the spectrometer; suppose that in fact 20 repeated analyses for iron are run on the same spectrometer with the same environmental conditions (temperature, humidity, etc., as well as the same operator using the standard methods). It is to be expected that the 20 resulting numbers will exhibit variability and that quite possibly none of the 20 would be exactly equal to μ, the true iron content. In fact, as mentioned above, the iron ppm reading that the spectrometer produces on any one of these repetitions is directly related to the number of iron atoms in the burning oil that are excited to the correct state to emit light at the particular iron frequency being monitored: from one to another of these 20 repetitions there will undoubtedly be variation in the actual number of iron atoms that are excited to the required degree.

A plausible physical explanation for this variability of excited atoms, for burns of fixed time, (see [1]) is as follows: at any given instant of time while the oil is burning, a large number N of distinct iron atoms is within the portion being burned; the ratio of N to the total number K of atoms burning at this instant is μ, the true iron ppm content. Each of the N atoms available either does or does not reach the required state to emit the particular spectral line to be monitored in the analysis; the

proportion of those available to reach this state which actually do reach this state is p. Furthermore, each individual iron atom either is or is not excited to the necessary state independently of all the other atoms. Then, as is well known, the number X to reach the necessary state at this given instant is a binomial random variable with parameters N and p. Since N is very large, then, as is also well known, X is essentially a normal random variable with mean Np and variance $Np(1-p)$ (the only two parameters in the distribution of X).

The actual iron reading which the spectrometer produces is directly related to an "average" over all the instants included in the fixed burning period and is "normalized" essentially by dividing by the total number of atoms, K, times the proportion p that should have been excited to the necessary state at any instant. Thus, the final spectrometer readout is essentially $\frac{X}{Kp}$, which is then approximately a normal random variable with mean $\frac{Np}{Kp} = \frac{N}{K} = \mu$ and with variance

$$\sigma^2 = \frac{Np(1-p)}{K^2p^2} = \frac{(1-p)}{Kp}\,\mu = \frac{(1-p)}{Np}\,\mu^2 .$$

We may thus conjecture that the variance in the spectrometer readout is a linear or quadratic function of the mean. In Section III, we give the results of an analysis of the Air Force data (from [5]) which appears to support an assumption of normality of sample readings with the variance being a quadratic function of the mean.

If the length of the source burn time is controlled by fixed

reference, rather than fixed time (that is, the source burn is terminated when the total energy received at a reference frequency, such as a carbon line, reaches a certain threshold level), a physical explanation of the variability of excited atoms may be given as follows: the number T of burning instants required until the threshold is reached with the reference line is random. If it is assumed that at each instant, independent of other instants, either the reference integrator receives an impulse (say, with probability ρ), or it does not (with probability $1 - \rho$), and if r impulses are required to reach the reference threshold, then T has a negative binomial distribution with parameters r and ρ. The energy accumulated at the iron line being monitored is thus $\sum_{i=1}^{T} X_i$, where, as before, X_i is the number of iron atoms reaching the necessary state in the i^{th} instant (so X_i is approximately normal with mean Np and variance Np(1-p)). Now if the spectrometer is properly calibrated, the readout $\hat{\mu} = \sum_{i=1}^{T} X_i$ has as its mean a value proportional to the true iron content μ. Since

$$E(\sum_{i=1}^{T} X_i) = E(T)E(X_i) = \left(r + \frac{r(1-\rho)}{\rho}\right) Np = \alpha\mu,$$

where α is a proportionality constant, we have $\rho = Npr/\alpha\mu$. The variance of $\hat{\mu}$ is

$$\begin{aligned} V(\hat{\mu}) &= E(T)V(Y_i) + V(T)E^2(Y_i) \\ &= r\left(1 + \frac{1-\rho}{\rho}\right) Npq + \frac{r(1-\rho)}{\rho^2} N^2p^2 \\ &= (1-p-Np)\ \alpha\mu = \frac{1}{r}\alpha^2\mu^2, \end{aligned}$$

using the above expression for ρ.

Thus in this case, as in the last, one may conjecture that the variance in readings is a quadratic function of the mean. Of course, confounded with the variance due to the physical process of energy emission are additional factors such as the effects of variation in calibration runs and variation due to other types of error such as those discussed in Section II. It is therefore of interest to test the hypothesis that such a relationship exists using actual experimental evidence (Section III).

If in fact the results of 20 repeated analyses were available, the average of the 20 readings should also be a normal random variable and standard techniques are available to make inferences about the unknown iron concentration μ in the crankcase sampled, given the 20 sample analyses. Of course, in practice more than one element is simultaneously analyzed during the same burn and typically 4 or 5 different elements are all of use in monitoring a given engine type. Thus, the sample results are used to make inferences about more than one type of contaminant; since the amounts of several different contaminants are simultaneously estimated, interrelationships between the readout amounts of iron and of copper, for example, are possible. Section III reports some interesting findings concerning such interrelationships.

In the present section, 20 repeated analyses of the same sample have been discussed merely to illustrate a plausible model to explain the inherent variability observed from one such analysis to another. It is not suggested that the current procedures should be modified to allow repeated spectrometric analyses of the same sample. Once

this inherent variability has been measured it is certainly possible to proceed with only a single analysis of each of the samples taken on a regular basis. Any conclusions derived about the probable amount of contaminant in the reservoir should be made with this variability well in mind.

III. SOME PARTICULAR RESULTS

1. Air Force Data

In this section some results derived from a study of data collected by the Air Force will be presented. These data were summarized in [5]; the authors would like to thank Mr. Donald C. Kittinger of WPAFB for making the original data collected available to us.

In 1967 the Air Force sent the same 190 oil samples, over a period of about one month, to each of 25 different laboratories to be analyzed on the spectrometers then used by these laboratories. The 190 samples were sent in different orders to the different labs and different numbering schemes were used to identify the samples, from one lab to another, so that the labs could not communicate with each other about specific readings they observed for the various samples. The purpose of the Air Force study was twofold: to see how consistently each given lab would get the same readings from the same oil sample, and to see how closely the results would agree from one laboratory to another. Unknown to the participating laboratories, the 190 samples consisted of 10 different samples, each repeated 10 times (making 100 samples

in all) plus 90 additional distinct samples, each sent only one time: thus, actually only 100 different (90 + 10) samples were used in the study. Of the 100 different samples of oil, 10 were standard mixtures with a known composition; the remaining 90 were merely selected from available used oil and were of unknown true composition. One of the 10 standard mixtures was repeated 10 times.

The Naval Air Rework Facility at Pensacola (NAVAIREWORKFACPENS), the laboratory which initiated the NOAP program, was one of the participants in this Air Force study. Since NAVAIREWORKFACPENS was the major supplier of data for the current contract, the original analyses they ran on the 10 sets of 10 repeated samples have been studied with great interest. Table 1 presents the sample means and standard deviations for each of the elements measured by Pensacola, as well as the sample sizes. Each sample should have occured 10 times, but some data is missing.

First this data has been used to test the hypothesis that the sample readings from the Pensacola spectrometer are normally distributed; this hypothesis is accepted with a significance level $\alpha = .05$ (see the appendix for the details of this test). Then, granting that the normal assumption is justified, it is of interest to investigate the interrelationships between the observed readings of the various elements, that is to test the hypotheses that the correlations between pairs of elements is zero. Because of a possibly rather complex relationship (mentioned above) between the true average reading μ for a given element and the variance of a single reading for the same element, it was felt that the covariance

Table 1

	Alum	Iron	Chrom	Silver	Copper	Tin	Magnes	Nickel	Silicon
Sample Group	1.11	75.56	7.22	3.11	61.44	8.33	2.89	1.89	7.11
1 (n=9)	1.17	2.83	3.63	0.93	2.30	4.12	0.33	1.83	0.78
Sample Group	0.20	145.70	12.20	2.60	14.20	12.20	11.30	3.30	11.70
2 (n=10)	0.42	7.85	3.22	0.84	1.14	3.19	0.82	0.67	0.48
Sample Group	0.50	82.10	8.90	4.10	6.90	5.60	15.20	1.20	3.70
3 (n=10)	0.71	3.38	2.02	0.57	1.10	2.50	1.03	0.63	0.48
Sample Group	23.12	27.75	11.75	4.50	34.88	28.00	10.37	13.50	3.12
4 (n=8)	1.73	0.71	2.71	0.53	1.13	3.55	0.52	0.53	0.35
Sample Group	3.50	136.20	10.60	4.80	108.40	10.40	5.50	1.00	11.40
5 (n=10)	1.27	5.12	5.58	1.03	5.97	4.62	0.53	0.47	0.52
Sample Group	0.50	109.70	9.90	2.80	13.70	9.70	19.40	2.90	7.20
6 (n=10)	0.71	7.53	2.18	0.63	1.57	5.85	1.43	0.74	0.63
Sample Group	3.11	100.22	9.22	4.00	82.78	10.33	26.89	1.67	11.67
7 (n=9)	1.54	6.04	2.28	0.87	5.65	4.61	1.17	1.32	0.50
Sample Group	0.40	131.40	5.40	3.20	6.60	6.40	11.60	2.50	5.20
8 (n=10)	0.70	5.70	3.34	1.03	1.17	4.72	0.70	1.08	0.63
Sample Group	1.29	17.57	9.86	1.86	10.29	5.43	12.29	0.71	4.71
9 (n=7)	0.95	0.79	2.41	0.69	0.76	2.51	0.49	0.49	0.49
Sample Group	0.33	198.56	8.00	2.33	8.78	10.00	10.11	4.33	3.56
10 (n=9)	0.50	11.29	2.06	0.71	1.20	5.10	0.60	1.12	0.53

Sample means (1st line) and sample standard deviations for 9 elements in 10 oil samples.

between pairs of elements might also depend upon the average contaminant level of the two elements involved. Thus, since the average level of contaminants varies widely from one to another of the 10 samples, correlations between elements were computed within each of the 10 samples and these were not pooled together. Table 2 gives the number of times (from the 10 different samples) the correlation between the various pairs was significant at level $\alpha = .05$. Note that particularly strong correlations seem to exist between pairs from the sets {copper, iron, magnesium} and {chromium, silver, nickel}. Thus it would appear that the readings on the various elements are not independent and that an erroneously high reading on copper, for example, may also bear some information about the error in the same analysis of the sample's content of iron and magnesium as well. This point will be touched on again in Section IV, in which we discuss a possible objective rule for identifying discrepant engines.

The latest Tri-Services recommendations on the required specifications for spectrometers to be used in oil analysis, and discussions with representatives of Baird-Atomic, Inc., the manufacturer of the machine at Pensacola, indicate that, for modern spectrometers, the variability in readings for any given element is dependent on the true average content of the element. (Some physical considerations on this point were discussed in Section II.) It was felt, therefore, that such a relationship might hold for the older Baird-Atomic machine at Pensacola. The Air Force data mentioned above was used to investigate such a possible relationship.

Table 2

	Al	Fe	Cr	Ag	Cu	Sn	Mg	Ni
Fe	0							
Cr	0	0						
Ag	1	2	6					
Cu	0	5(1)	2	1(1)				
Sn	0	1(1)	1	1	(1)			
Mg	0	5	(1)	(1)	4	0		
Ni	(1)	1	4	2	1	1	0	
Si	1	1	1	1	2	(1)	0	2

Numbers of significant correlations

2-tailed test, $\alpha = .05$

( ) indicates negative correlation

Specifically, for modern machines, the relationship is assumed to be

$$\sigma^2 = a + b\,\mu^2,$$

where μ is the true ppm content of a given element in the oil, σ^2 is the variance of repeated analyses of the same sample (for the same element) and a and b are constants. Note that this assumed relation is in agreement with those presented in Section II. With the 10 samples of approximately 10 analyses each, then, it was possible to estimate μ and σ^2 for each element within each sample; for a given element, such as iron, let $\bar{X}_i$, $i = 1,2,\ldots,10$ and S_i^2, $i = 1,2,\ldots,10$, denote the estimates of the corresponding μ_i and σ_i^2, respectively. Then, again, for each element, the coefficients a and b in the equation $S_i^2 = a + b\,\bar{X}_i^2 + e_i$, $i = 1,2,\ldots,10$ can be estimated from the observed data using standard regression theory and, assuming that the observed deviations about this straight line are normally distributed, the hypothesis that $b = 0$ can be tested for each element. (See Table 3 in the appendix.) Of course, if the hypothesis $b = 0$ is accepted, then there is some evidence that the variance in individual readings for the given element does not depend on the actual content of the element over the range of contents covered; if it does not appear from the data that $b = 0$, then there is some evidence that in fact the older machine currently in use exhibits a relation between σ^2 and μ similar to that of modern machines. Using this procedure with the Air Force data, it was found that for aluminum, iron, copper and magnesium the

coefficient b is significantly greater than 0, with a test of size $\alpha = .01$; for the other elements analyzed b does not differ from 0 even with $\alpha = .20$. It should be mentioned that the true content of these other elements apparently did not vary much from sample to sample. A similar analysis could be used to investigate possible relationships between the covariance of any pair of elements and the average content of each element, but lack of time has precluded such an investigation at this time. As can be noted in Table 1, the apparent content of iron and also of copper in the 10 samples goes well beyond the practical limits observed in NOAP. Thus, for these two elements, the relationships between the variance and the average content may not be as notable when the range of content represented is more realistic of that found in operating engines.

Three possible conclusions seem justified from this study of the repeated samples run by NAVAIREWORKFACPENS:

(a) The readings for any given element do appear to be normal random variables.

(b) The readings of several pairs of elements do not seem to be independent and objective rules for determining discrepant engines should allow for this possibility.

(c) It appears that the variances of readings made on the Pensacola machine are not independent of the actual concentrations. However, this point should be investigated more thoroughly by running a well-planned set of analyses on the Pensacola machine with realistic levels on all the elements analyzed.

The covariance structure between pairs of elements should also be investigated. It is possible that for realistic levels of the ppm content it can safely be assumed that the variances of individual readings of a given element are essentially constant. Reasonable estimators for μ and σ^2, under the assumption that $\sigma^2 = a + b\,\mu^2$, are discussed in the appendix.

2. NOAP Data

NAVAIREWORKFACPENS has provided a data tape containing records of all the operational analyses they performed during a three month segment of time from July 1, 1967 to September 30, 1967. The authors would like to thank Mr. B. B. Bond, NAVAIREWORKFACPENS, for making this data available. Roughly 21,000 separate oil analyses are included; for each analysis the particular model number and serial number of the item samples are listed, as well as the date the analysis was performed, the number of hours since overhaul and the number of hours since oil change, and the ppm readings of each of 10 elements: aluminum, iron, chromium, silver, copper, tin, magnesium, lead, nickel and silicon. The tape contains no information about any action the lab may have recommended on the basis of a given analysis, nor, if action were taken, whether the lab recommendations proved accurate. The model number designates the type of aircraft engine (or gear box or transmission or whatever) which was sampled from, while different serial numbers identify different particular units of the given type.

The tape was first searched to identify the different model numbers represented in the 21,000 analyses and the different serial

numbers within each model number, as well as the number of times each separate serial number occurred. Then, the most frequently occurring model number (R182082, a Wright reciprocating engine) was selected for investigation, since it would provide the largest possible amount of data. Some 600 different analyses from this model (with no control on the different serial numbers involved) were plotted by the computer; for each element, the computer plotted the ppm content versus the number of hours since oil change. It was expected that at least some of the elements would show a buildup in amount as hours since oil change increased. For iron, copper and aluminum (see Figures 1, 2 and 3) this does seem to be the case, while the other seven elements evidenced no distinct trend in corresponding plots of 600 analyses.

Then, to further investigate possible buildups in content as hours since oil change increased, five particular serial numbers were selected from all those available for this model. For each of the five serial numbers, for each element, the computer plotted the ppm count versus hours since oil change for all analyses available during this three month period. Figures 4, 5 and 6 show these plots for iron, copper and aluminum. Five different symbols are used, X, +, Δ, □, ◇, to represent the five serial numbers. Thus it is possible from these plots to see the buildup, if any, of the particular element involved for each serial number, making it easy to graphically compare different serial numbers of the same model. The other seven elements showed no clear evidence of a consistent trend, for this model number, so their plots are not

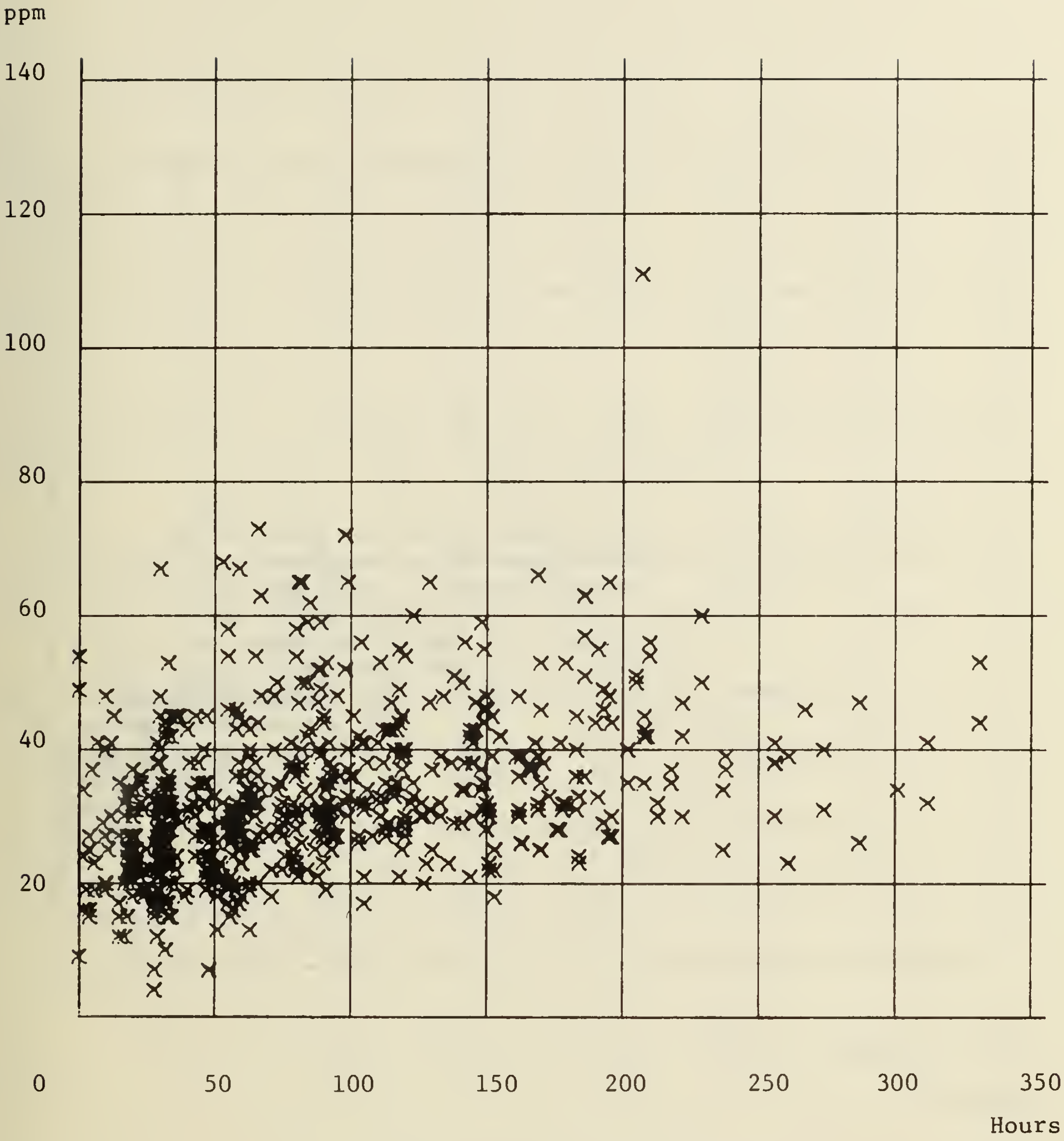

Figure 1

Hours since oil change vs ppm iron.

(Model No. R182082)

[illegible]

[illegible]

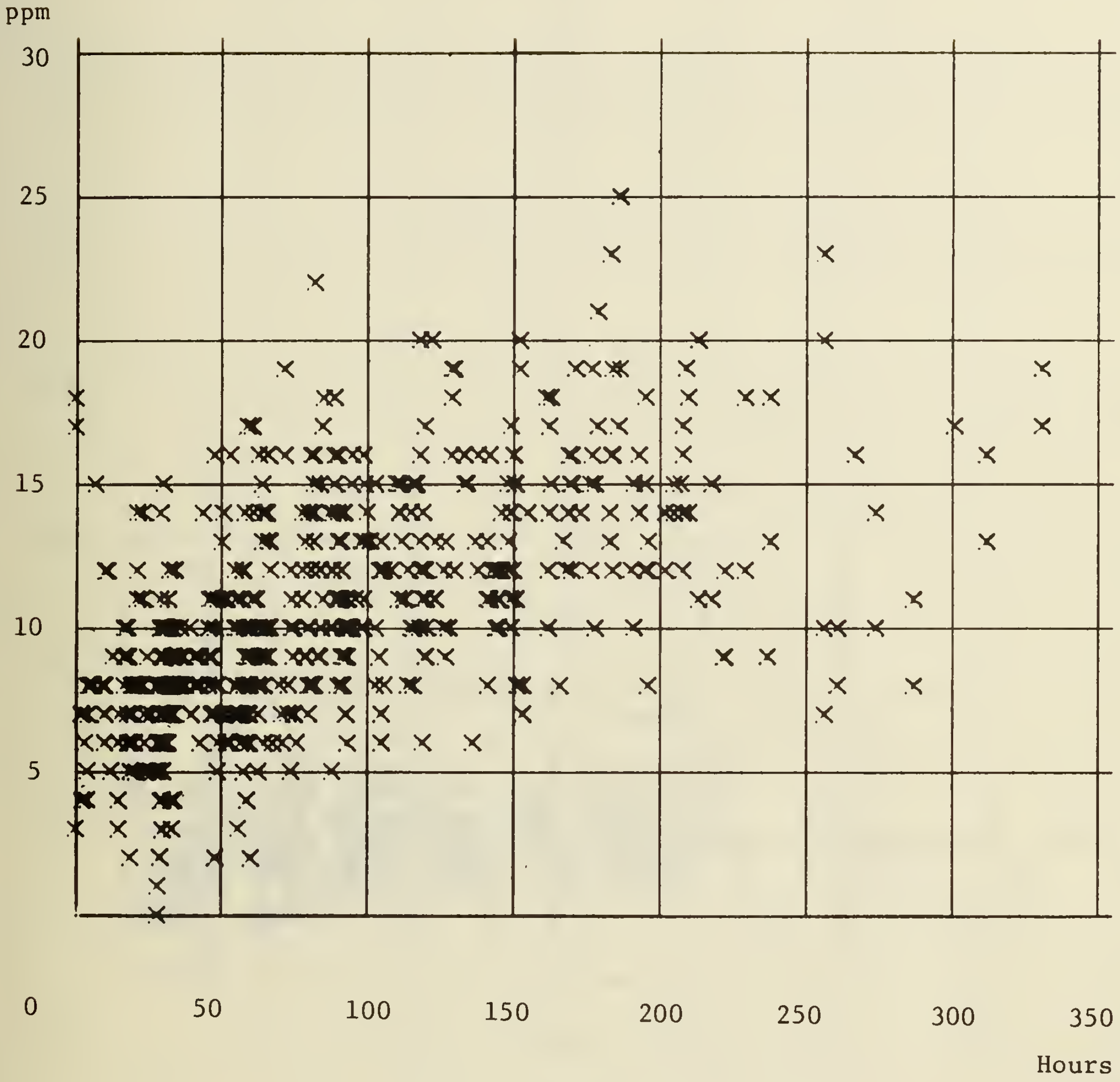

Figure 2

Hours since oil change vs ppm copper.

(Model No. R182082)

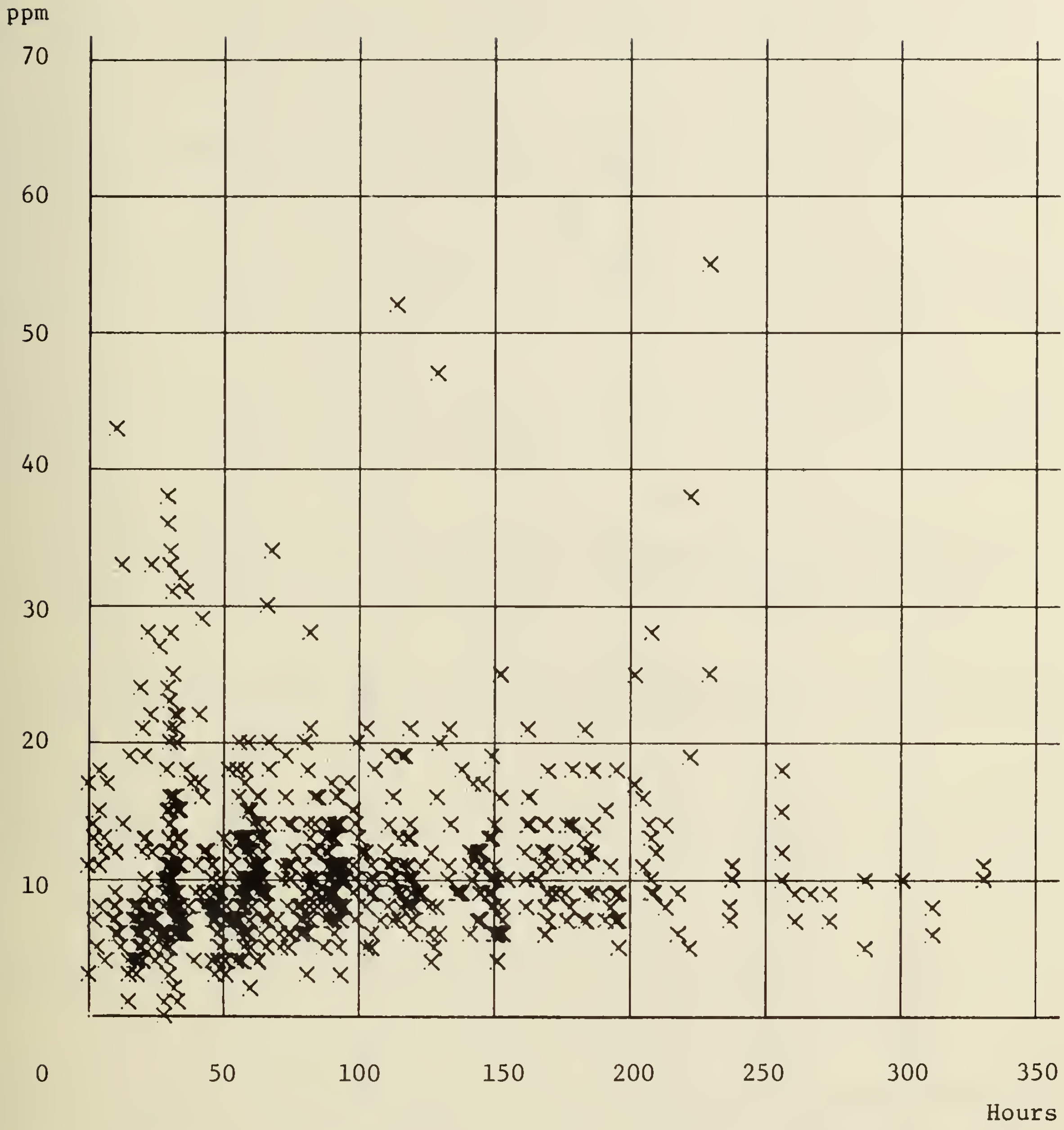

Figure 3

Hours since oil change vs ppm aluminum.

(Model No. R182082)

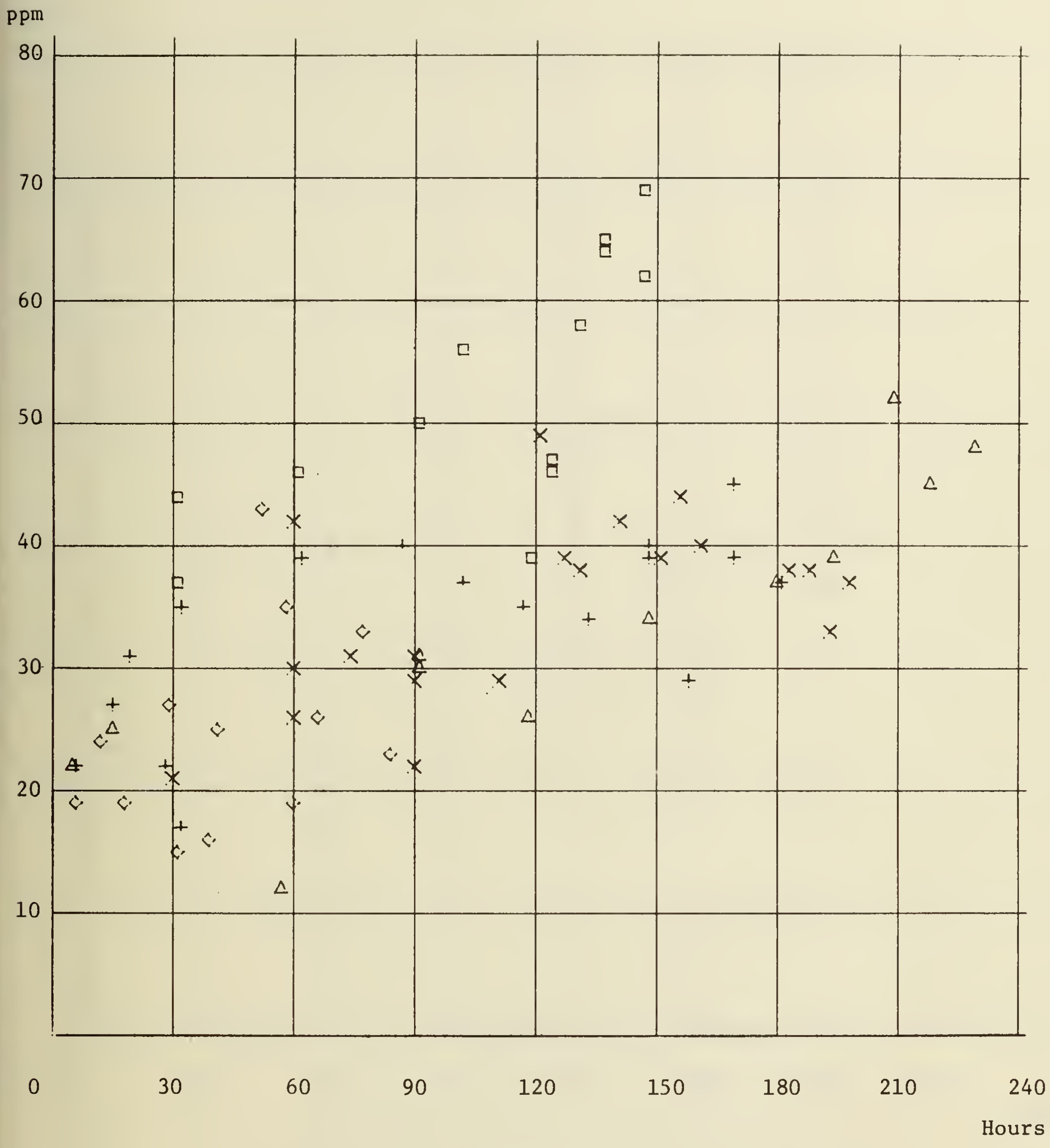

Figure 4

Hours since oil change vs ppm of iron.

(Different serial nos. of Model No. R182082)

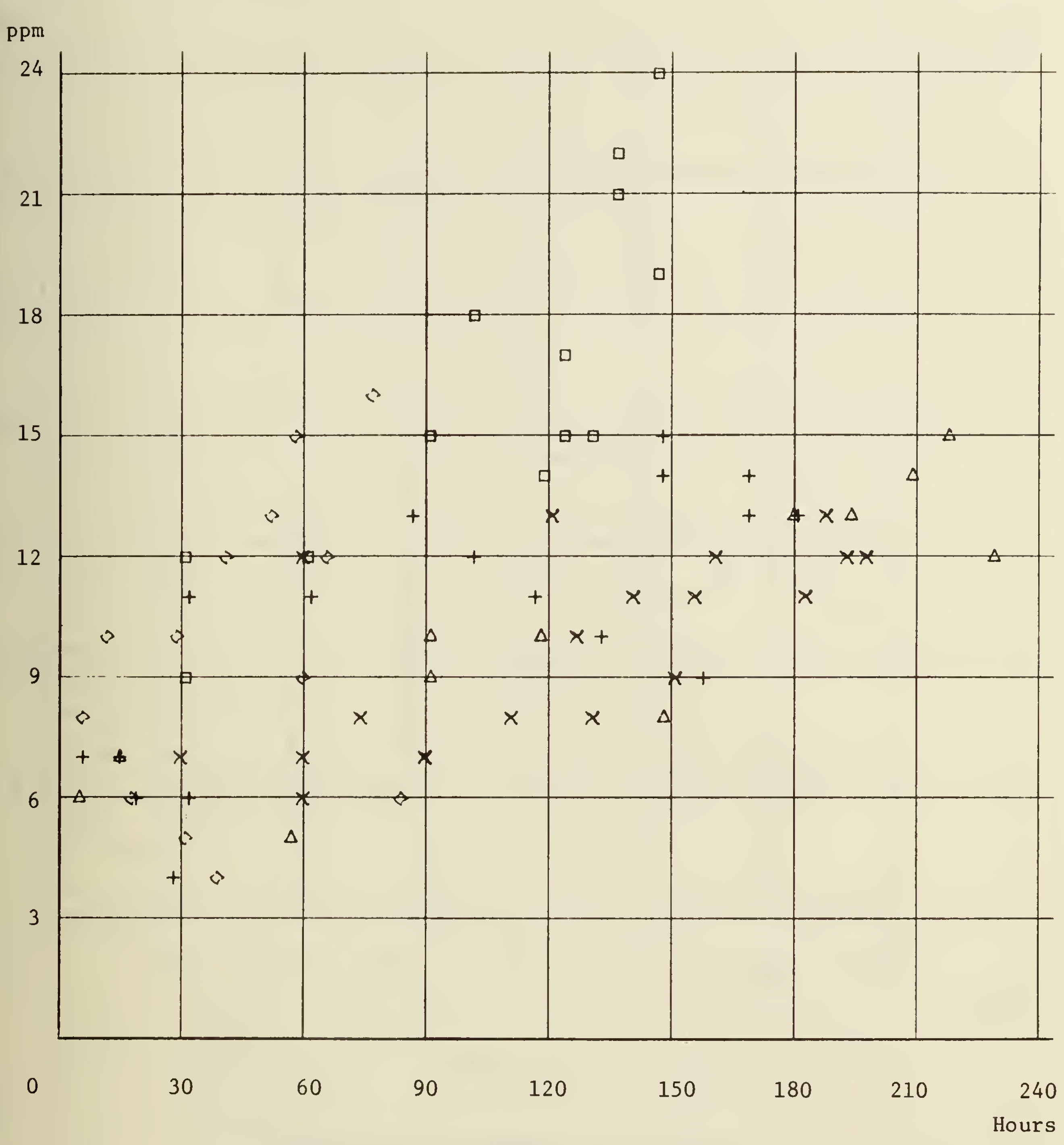

Figure 5

Hours since oil change vs ppm copper.

(Different serial nos. of Model No. R182082)

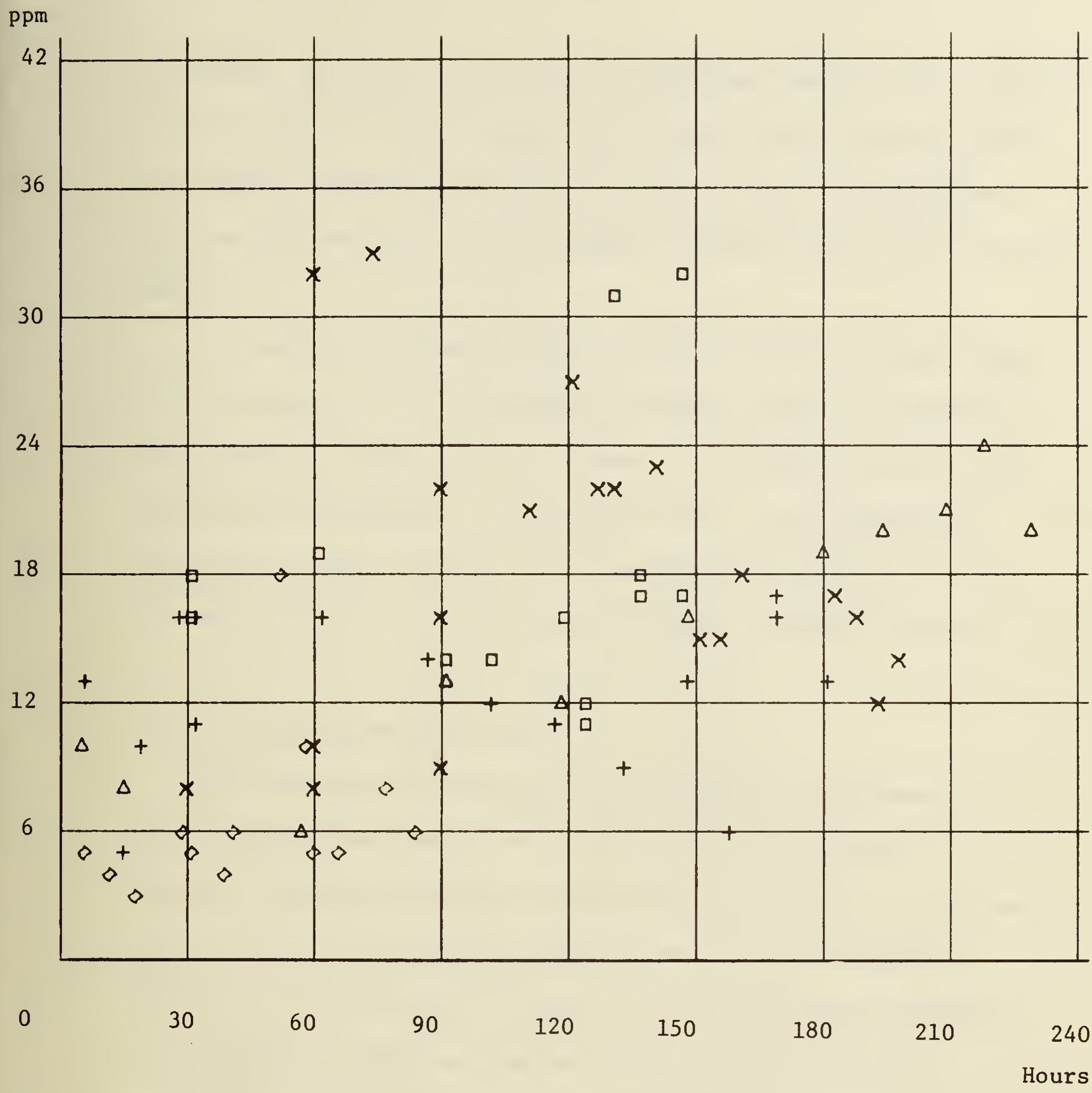

Figure 6

Hours since oil change vs ppm aluminum.

(Different serial nos. of Model No. R182082)

presented. Note that for each of these three elements there appears to be a roughly linear increase in the ppm content as hours since oil change increase for each of the five aircraft. Furthermore, this buildup appears to be at roughly the same rate for each serial number.

It would seem possible that a general buildup in content might also occur as hours since overhaul increase, given an essentially fixed number of hours since oil change. This point has been only superficially examined at this time; however, this superficial examination seems to indicate no consistent trend as hours since overhaul increase for any element, for these particular aircraft.

IV. A SUGGESTED OBJECTIVE RULE

The preceding sections have been devoted to a discussion of the current methods now in use in NOAP, the possible errors in the spectrometer ppm readings and some particular results discovered from a study of the Air Force data and of the actual analysis records of a 3 month period of time. In this section a procedure for identifying discrepant engines will be discussed which specifically allows, and takes advantage of, the particular phenomena mentioned in Section III.

It seems clear that many different types of failure cannot be detected by spectrometric oil analysis. For example, a failure that occurs as a discrete event, such as the sudden collapse of a bearing, would quite possibly not be preceded by an unusual wearing mechanism which deposits unusual quantities of the bearing metal

in the oil reservoir. Thus, it is not expected that such catastrophic events can be detected or predicted from spectrometric oil analysis. At the same time, the success of NOAP testifies to the existence of many types of failures which can be detected by engine oil analysis.

Those failures which can be detected are the ones which are associated with an abnormally high metallic content in the oil prior to their occurrence (for a sufficiently long period of time to permit a good likelihood that a high content sample is taken). Thus, any objective rule for detecting discrepant engines should be one which identifies abnormally high contents of one or more elements. Figures 1, 2 and 3 in Section III make it seem possible that the content which is called abnormally high may be dependent on the number of hours since oil change (at least for model R182082). That is, granted that these figures indicate that the typical or normal content seems to increase with hours since oil change, then it seems logical that a reading that is high for 8 hours after oil change may well be normal or typical for 20 hours after oil change since the average content is higher at the later time. Thus, the limits defining excessively high content of any particular metal might also be expected to increase with hours since oil change.

In addition, since the variances in readings for some elements are apparently a function of the mean concentration, it is possible that the variance-covariance structure is dependent on time since oil change. Since the variance appears to increase with increasing mean, which in turn tends to increase with time since oil change,

the net effect could be to increase even more the limits defining excessively high content. We have not incorporated the latter effect in the suggested objective rule, however, since the actual magnitude of increase in mean concentration with time since oil change in small, which appears to make the amount of change in the variance-covariance structure with time since oil change negligible. (In this connection, see Table 3 in the appendix for estimates of b in the relationship $\sigma^2 = a + b\,\mu^2$.)

Assuming that the true ppm content μ, for any particular element within any particular aircraft, is linearly increasing with time since oil change, standard statistical techniques are available for estimating μ from sample data, as well as for identifying those particular readings which seem excessively high. Readings which seem excessively high, of course, might be expected from discrepant engines, whose true content has increased at a faster rate than the typical or to a higher value than typical. Since the spectrometer simultaneously analyzes for several different metallic contaminants and, as noted in Section III, the readings are correlated between some of the elements, an efficient procedure should make use of all the information possible about any given element, including the correlations with other readings. The technique which seems ideally suited for describing the behavior of normal content and for identifying abnormally high content at any sampling point is least squares or regression analysis.

Briefly, this method and its suggested use may be described as follows. All of the different serial number engines of the same unit

model number are almost identical in makeup. It might then be expected that the normal buildup of contaminants in a particular engine would be essentially the same as for any other of the same type. (A very preliminary analysis of different serial numbers seems to deny this, but more investigation is necessary before a reliable conclusion can be made.) If all engines of the same unit number do have essentially the same normal concentration buildup, then data from all such engines can be combined and used to estimate the normal trend (as time sence oil change increases) of each contaminant for all these engines. If it is determined that the different engines of a given type do not have essentially identical patterns of buildup, then the data for any given engine should be used to estimate normal buildup for only that engine. The point to be stressed here is whether or not data can be pooled for all engines of the same unit model; the suggested technique will be the same in either case, but the accumulation of data and thus the accuracy of the procedure will be greatest and quickest if it is valid to pool data for all engines.

As has been stressed, the accumulation of some or all of the 10 elements analyzed may be of interest for any given aircraft. The true accumulation for all 10 elements then is a vector $\underline{\mu}$ having 10 components, one for each element. As operating time passes, the true accumulation vector takes on different vector values. In order to stress the possible dependence on hours since oil change (hours since overhaul can be handled in a similar manner if it proves of use), let $\underline{\mu}(t)$ represent the true accumulation at t hours since

oil change. Then (at least for model R182082) the true behavior of $\underline{\mu}(t)$, for any given aircraft, seems to be fairly well approximated by

$$\underline{\mu}(t) = \underline{a} + \underline{b}\, t, \tag{1}$$

where $\underline{a}$ and $\underline{b}$ are each 10 x 1 vectors and t is the (scalar) hours since oil change. Thus, for example, the $i^{\underline{th}}$ component of $\underline{a}$ gives the amount of the $i^{\underline{th}}$ contaminant to be expected immediately after oil change and b_i gives the rate of accumulation of the $i^{\underline{th}}$ contaminant per hour, $i = 1,2,\ldots,10$. It is quite possible for b_i to be zero for one or more elements, that is, for the amount of any particular contaminant to remain essentially the same, no matter how many hours have passed since oil change.

Assume, then, that the oil of a given aircraft has been sampled at each of n times (hours since oil change) $t_1, t_2, \ldots, t_n$, and that each such sample has been analyzed on the spectrometer and that $\underline{Y}(t_1), \underline{Y}(t_2), \ldots, \underline{Y}(t_n)$ are the n 10 x 1 vectors of readings from the n samples. As has been mentioned earlier, it seems reasonable that $\underline{Y}(t_i)$ is a multivariate normal vector with mean $\underline{\mu}(t_i)$ and a possibly non diagonal covariance matrix Σ. The components of Σ will consist of two distinct parts. First, as noted in Section III, repeated readings on the same sample seem to be correlated and these will affect the off-diagonal components of Σ. Second, equation (1) expresses a linear assumption about the true content as hours since oil change increase. Inadequacies of this assumption (deviations from linearity) may affect both diagonal and off-diagonal elements of Σ. Also, as noted above and in Sections II and III,

it appears possible that the variance of readings of any given element is related to the true content of the element in the oil. Thus as the true content increases it would be expected that the variances of the readings would also increase. However, the plots examined show a relatively slow buildup for normal engines and it is anticipated that the variances of the readings will shift by a negligible amount; thus, it seems safe as a first approximation to assume that $\not{Z}$ remains constant and does not change as hours since oil change increases.

If $\not{Z}$ were known, then straightforward weighted least squares could be used to estimate $\underline{a}$ and $\underline{b}$, given a set of sample readings. Since $\not{Z}$ is not known, it must be estimated from sample data for each given engine (or engine type). The estimate $\hat{\not{Z}}$, along with estimates $\hat{\underline{a}}$ and $\hat{\underline{b}}$ of $\underline{a}$ and $\underline{b}$, can then be used to construct a good objective rule. Details on how a set of sample readings can be used to get estimates $\hat{\underline{a}}$, $\hat{\underline{b}}$ and $\hat{\not{Z}}$, of $\underline{a}$, $\underline{b}$ and $\not{Z}$, respectively, are given in the appendix.

Once estimates $\hat{\underline{a}}$, $\hat{\underline{b}}$ and $\hat{\not{Z}}$ are available for a given engine, they can then be used to define a 10-dimensional region $R_\alpha(t)$ for any number t of hours since overhaul with the following property: given a sample from a normal engine at t hours since oil change, one whose increase in content has followed its own previous normal history, the probability is approximately $1 - \alpha$ that the vector $\underline{Y}(t)$ falls within $R_\alpha(t)$ and the probability is approximately α that it does not fall within $R_\alpha(t)$. The parameter α may be set at any desired level, say .05, .01 or .001. Then if a sample

taken at time t results in an analysis vector for this given engine which happens to fall outside $R_{\alpha}(t)$, either a relatively rare event has occurred (given the engine is normal), or the true ppm content of the engine at the given time is in excess of a normal amount for one or more elements or combination of elements. Thus, the suggested objective rule for identifying discrepant engines is: *Use all previous data for the given engine to estimate* $\underline{a}$, $\underline{b}$ *and* $\underline{\Sigma}$. *Determine* $R_{\alpha}(t)$ *for the given value of* t *of the incoming current sample. If* $\underline{Y}(t)$, *the current analysis, falls outside* $R_{\alpha}(t)$, *call the engine discrepant and take appropriate action.* The details of computation of these quantities are given in the appendix.

It should be pointed out that a procedure more similar to the one currently used could easily be defined by using two or more values of α. For example, one might want to sample the engine more frequently if a fairly rare event has occurred and actually recommended grounding the aircraft only if a very rare event has occurred. This could be accomplished as follows: choose $\alpha_1 = .20$ (for example) and $\alpha_2 = .01$ (for example). Then, if the sample analysis vector $\underline{Y}(t)$ falls in $R_{\alpha_1}(t)$ do nothing; if $\underline{Y}(t)$ falls outside $R_{\alpha_1}(t)$ but inside $R_{\alpha_2}(t)$ then sample at a greater frequency; if $\underline{Y}(t)$ falls outside $R_{\alpha_2}(t)$ then ground the aircraft. In using a procedure of this type, one essentially has control over how often one type of error may occur. That is, since $\underline{Y}(t)$ would be outside $R_{\alpha_1}(t)$ and inside $R_{\alpha_2}(t)$ with probability p_1 between .01 and .2, if the engine is normal, more frequent sampling than normal would occur the proportion p_1 of the time when it wasn't

needed. Similarly, since $\underline{Y}(t)$ would be outside $R_{\alpha_2}(t)$ with probability $p_2 = .01$ if the engine is normal, the proportion p_2 of normal engines would be needlessly grounded. By adjusting α_1 and α_2 these two risks may be made as large or as small as is desired.

A second type of error may also occur, namely, a plane which should have been grounded may not actually be grounded. It is very difficult to estimate the actual probability β of this error occurring for a given α, but it can be shown in general that the larger α is taken, the smaller β will be, and vice versa. Furthermore, under a fairly wide range of conditions the objective rule we are proposing can be expected to have the smallest possible β for any given value of α.

V. APPENDIX

1. Estimation of $\underline{a}$, $\underline{b}$ and $\not{\Sigma}$

Given $Y(t_i)$; $i = 1,2,\ldots,n$, is a sample of n independent 10×1 vector observations and that $\underline{Y}(t_i)$ is multivariate normal with mean $\underline{\mu}(t_i) = \underline{a} + \underline{b}\, t_i$ and variance-covariance matirx $\not{\Sigma}$; $i = 1,2,\ldots,n$, define the $10 \times n$ matrix Y by $Y = (\underline{Y}(t_1),\ldots,\underline{Y}(t_n))$. Let X denote the $2 \times n$ matrix

$$X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_n \end{pmatrix} = (\underline{X}_1,\ldots,\underline{X}_n).$$

and β the 10×2 matrix $\beta = (\underline{a},\underline{b})$. Since $\underline{Y}(t_i) \sim N(\beta\underline{X}_i, \not{\Sigma})$, our model for the n analysis vectors can be written

$$Y = \beta X + e$$

where $e = (\underline{e}_1,\ldots,\underline{e}_n)$ is a $10 \times n$ matrix whose columns are independent multivariate normal random vectors with zero mean and variance-covariance matrix $\not{\Sigma}$. As is shown in Anderson [2], for this model the maximum likelihood estimator for β is given by

$$\hat{\beta} = \underline{Y}\, X'(XX')^{-1} = (\hat{\underline{a}},\hat{\underline{b}}),$$

independent of $\not{\Sigma}$, where ' denotes transpose. This estimator is the minimum variance linear unbiased estimator for β. The maximum likelihood estimator for $\not{\Sigma}$ is given by

$$\hat{\not{\Sigma}} = \frac{1}{n}(Y-\hat{\beta}X)(Y-\hat{\beta}X)'$$

and $S^2 = \frac{n}{n-2}\hat{\not{\Sigma}}$ is an unbiased estimator for $\not{\Sigma}$.

2. Construction of $R_\alpha(t)$

As was discussed in Section IV, $R_\alpha(t)$ is to be a region such that the probability is at least $1 - \alpha$ that $\underline{Y}(t)$ belongs to $R_\alpha(t)$, for any number t of hours since overhaul. Given estimates $\hat{\underline{a}}$, $\hat{\underline{b}}$ and $\hat{\not{\Sigma}}$, the vector

$$\hat{\underline{\mu}}(t) = \hat{\underline{a}} + \hat{b}\, t = \hat{\beta}\underline{T} = \sum_{i=1}^{n} \underline{Y}(t_i)\underline{X}_i'(XX')^{-1}\underline{T},$$

where $\underline{T}' = (1,t)$, is an estimate of the true mean content $\underline{\mu}(t) = \underline{a} + \underline{b}\, t$. The variance-covariance matrix of $\hat{\underline{\mu}}(t)$ is easily obtained as follows: noting in equation (2) that $\hat{\beta}\underline{T}$ is a linear combination of the independent vectors $\underline{Y}(t_i)$, it follows that (see [2] or [4])

$$\Sigma_{\hat{\underline{\mu}}(t)} = \sum_{i=1}^{n} [\underline{X}_i'(XX')^{-1}\underline{T}]^2 \Sigma$$

$$= \Sigma\, \underline{T}'(XX')^{-1}\underline{T}$$

The actual observed vector $\underline{Y}(t)$ is, of course, the sum of $\underline{\mu}(t)$, the true mean vector, plus the 10x1 observational error $\underline{e}$. Since $\underline{Y}(t)$ and the columns of Y are independent, the variance-covariance matrix of the difference $(\underline{Y}(t)-\hat{\beta}\underline{T})$ is $\Sigma + \Sigma_{\hat{\mu}(t)} = (1+\underline{T}'(XX')^{-1}\underline{T})\Sigma$. It follows that

$$\frac{(\underline{Y}(t)-\hat{\beta}\underline{T})}{\sqrt{1 + \underline{T}'(XX')^{-1}\underline{T}}}$$

has a multivariate normal $(\underline{0},\Sigma)$ distribution, so

$$\frac{(\underline{Y}(t)-\hat{\beta}\underline{T})'}{\sqrt{1 + \underline{T}'(XX')^{-1}\underline{T}}}\,(n\hat{\Sigma})^{-1}\,\frac{(\underline{Y}(t)-\hat{\beta}\underline{T})}{\sqrt{1 + \underline{T}'(XX')^{-1}\underline{T}}}$$

has Hotellings T^2 distribution, and

$$H(\underline{Y}(t),Y) = \frac{(n-11)(\underline{Y}(t)-\hat{\beta}\underline{T})'[(Y-\hat{\beta}X)(Y-\hat{\beta}X)']^{-1}(\underline{Y}(t)-\hat{\beta}\underline{T})}{10(1+\underline{T}'(XX')^{-1}\underline{T}+1)}$$

has an F distribution with 10 and n - 11 degrees of freedom. It should be noted that these distribution results require $n \geq 12$.

Now for fixed $\underline{T}$, the probability is $1 - \alpha$ that $H(\underline{Y}(t),Y) \leq F(\alpha)$, where $F(\alpha)$ is the $100(1-\alpha)\underline{th}$ percentile of the tabulated $F_{10,n-11}$ distribution. For fixed t, define $R_\alpha^*(t)$ to be the set

$$R_\alpha^*(t) = \{\underline{y}(t) : H(\underline{y}(t),Y) \leq F(\alpha)\}. \qquad (1)$$

Then

$$P[\underline{Y}(t) \varepsilon R_\alpha(t)] = P[H(\underline{Y}(t),Y) \leq F(\alpha)] = 1 - \alpha$$

so $R_\alpha^*(t)$ is a $100(1-\alpha)\%$ confidence ellipsoid for $\underline{Y}(t)$, the observed vector of sample analysis results at t hours since oil change. However, from a consideration of the particular application we wish to make, assuming that only unusually <u>high</u> concentrations are indicative of trouble, it is suggested that $R_\alpha(t)$ should include all points in the set

$$\{\underline{y}(t) : \underline{y}(t) < \hat{\mu}(t)\} \qquad (2)$$

(meaning component-wise inequality). Thus the region $R_\alpha(t)$ is defined to be the union of the sets in (1) and (2),

$$R_\alpha(t) = \{\underline{y}(t) : H(\underline{y}(t),Y) \leq F(\alpha) \quad \text{or} \quad \underline{y}(t) < \hat{\mu}(t)\}.$$

The probability that $\underline{Y}(t)$ falls outside $R_\alpha(t)$ is thus strictly less than α. How much the actual probability differs from α is not known at the present time, but an evaluation of this difference should not prove to be an insurmountable problem. Using the set of points satisfying (1) or (2) thus provides a conservative region $R_\alpha(t)$; it seems quite feasible to evaluate how conservative it is and to find the exact probability α' that $R_\alpha(t)$ contains $\underline{Y}(t)$.

3. Testing the Normality Hypothesis

A test of the hypothesis that the observations from oil analyses may be considered to be drawn from normal populations may be performed using the data from the Pensacola lab in the Air Force experiment. Since the observations within a sample group (that is,

a group of analyses on the same batch of oil) appear to be correlated from one element to another, the following test procedure was used: For each group, the sample covariance matrix $\hat{\Sigma}$ was calculated, and a non-singular matrix P was found such that $P \hat{\Sigma} P' = I_9$. Thus, if the 9 x 1 vectors $\underline{X}_i$ of readings for the 9 elements in a given sample group were distributed $N(\underline{\mu},\hat{\Sigma})$, it would follow that $P\underline{X}_i \sim N(P\underline{\mu},I_9)$. Thus the components of the vectors $P(\underline{X}_i-\bar{\underline{X}})$ should be independent standard normal random variables.

Such a transformation P was found for each of the 10 sample groups, and the components of the resulting sample vectors were tested for normality using the Kolmogorov-Smirnov goodness of fit test. This procedure yielded a pooled sample size on the order of 900 (roughly, 9 elements x 10 sample groups x 10 observations per group). The test statistic D_n in the Kolmogorov-Smirnov test is given in this case by

$$D_n = \sup_x |F_n(x) - \Phi(x)| ,$$

where Φ is the standard normal distribution function and $F_n(x) = j/n$ for $X_{(j)} \leq x < X_{(j+1)}$ $(j = 0,\ldots,n)$, where in turn $X_{(k)}$ denotes the $k\underline{^{th}}$ largest value in the pooled sample of size $n(n \approx 900)$. The test indicates rejection of the hypothesis that the transformed observations are standard normal provided the observed value of D_n is sufficiently large. For the data mentioned above, the observed value of D_n is .033, which is not significant at the .05 level with a sample of the present size. That is to say, the test we are using will lead to rejection of the hypothesis of

normality, when in fact the data are from a normal population, with probability not more than .05. This outcome on the Kolmogorov-Smirvon test may be considered to be a strong evidence in support of the basic assumption that the spectrometer readings for the 9 elements monitored may be considered to be drawn from a multivariate normal population.

4. Estimation of μ_i and $\sigma_i^2 = a + b\,\mu_i^2$

Suppose X_{ij} is the $j^{\underline{th}}$ observation from the $i^{\underline{th}}$ sample group for a given element. In our model we may assume that

$$X_{ij} \sim N(\mu_i, a+b\mu_i^2); \quad i = 1,2,\ldots,10, \quad j = 1,2,\ldots,n_i(\approx 10)$$

where the X_{ij}'s are independent and the parameters a and b depend only upon the element involved. It is desired to find estimators $\hat{\mu}_1,\ldots,\hat{\mu}_{10}$, $\hat{a}$ and $\hat{b}$, for the parameters $\mu_1,\ldots,\mu_{10}$, a and b, with "good" properties. An effort directed toward finding the maximum likelihood estimators in this case yielded a system of nonlinear equations which we have not yet succeeded in solving in closed form, although in each particular case a numerical solution could be obtained. A reasonable alternative method which should give very nearly the best estimators is as follows: first, estimate each sample group mean μ_i by the corresponding observed sample mean,

$$\hat{\mu}_i = \bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i; \quad i = 1,2,\ldots,10.$$

Next, for each sample group compute the sample variance,

$$s_i^2 = \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2/(n_i-1)$$

Estimate a and b by least squares using standard linear regression theory with the 10 observed pairs of points $(s_1^2,\bar{x}_1^2),\ldots,(s_{10}^2,\bar{x}_{10}^2)$. This gives

$$\hat{b} = \frac{\sum_{i=1}^{10} (\bar{x}_i^2-\overline{\bar{x}^2})(s_i^2-\overline{s^2})}{\sum_{i=1}^{10} (\bar{x}_i^2-\overline{\bar{x}^2})} ,$$

$$\hat{a} = \overline{s^2} - \hat{b}\,\overline{\bar{x}^2} ,$$

where $\overline{s^2} = \sum_{i=1}^{10} s_i^2/10$ and $\overline{\bar{x}^2} = \sum_{i=1}^{10} x_i^2/10$. Finally, take

$$\hat{\sigma}_i^2 = \hat{a} + \hat{b}\,\bar{x}_i^2; \quad i = 1,2,\ldots,10 .$$

5. Testing whether b = 0 in the linear regression $S^2 = a + b\bar{x}^2$

In order to determine whether, for each element, the time since oil change is of significant value in making decisions concerning whether a concentration readout from the spectrometer indicates a discrepant engine, it is useful to test the hypothesis that b is zero. For, if the slope b (for a given element) in the linear regression equation $\sigma^2 = a + b\mu^2$ is zero, then the concentration readings from the spectrometer for that element in a given engine do not depend upon the buildup in mean concentration, or in turn, the time since oil change.

Let $\bar{x}_i$ and s_i^2 denote the observed sample mean and sample variance of the readings from $i^{\underline{th}}$ sample group for a given

element; $i = 1,2,\ldots,10$. The value of a and b may be estimated by $\hat{a}$ and $\hat{b}$ as discussed in the preceding section. In addition, the variance of the estimator $\hat{b}$ may be estimated by

$$s_{\hat{b}}^2 = \frac{1}{8}\,\frac{\sum_{i=1}^{10} (s_i^2-\overline{s^2})^2 - \hat{b}\sum_{i=1}^{10} (\bar{x}_i^2-\overline{\bar{x}^2})(s_i^2-\overline{s^2})}{\sum_{i=1}^{10} (\bar{x}_i^2-\overline{\bar{x}^2})^2}$$

Under the present assumptions, it follows that the quotient

$$T = \frac{\hat{b} - b}{S_{\hat{b}}}$$

has a t - distribution with 8 degrees of freedom. The hypothesis that $b = 0$ may be rejected if the calculated value of T is sufficiently large. A test which leads to an erroneous rejection of the hypothes that $b = 0$ with probability $\alpha = .01$, when in fact this slope is zero, is thus obtained by rejecting the hypothesis if the calculated value of T exceeds 2.75 (a one-sided size .01 t-test). The results of such tests calculated using the Air Force data from the Pensacola lab for the 9 elements monitored in that experiment, are summarized in Table 3. Note that, based upon these experimental results, there is apparently no significant dependence upon time since oil change for the elements chromium, silver, tin, nickel and silicon.

element	b	s_b^2	T	reject the hypothes that $b = 0$?
aluminum	$4.016\text{x}10^{-3}$	$1.823\text{x}10^{-6}$	3.0	yes
iron	$3.080\text{x}10^{-3}$	$8.305\text{x}10^{-8}$	10.	yes
chromium	$2.646\text{x}10^{-2}$	$8.870\text{x}10^{-4}$	.87	no
silver	$7.357\text{x}10^{-3}$	$5.800\text{x}10^{-5}$	.97	no
copper	$3.238\text{x}10^{-3}$	$1.054\text{x}10^{-7}$	32.	yes
tin	$-6.292\text{x}10^{-3}$	$1.288\text{x}10^{-4}$	-.17	no
magnesium	$2.224\text{x}10^{-3}$	$2.511\text{x}10^{-7}$	4.4	yes
nickel	$-4.177\text{x}10^{-3}$	$3.126\text{x}10^{-5}$	-.75	no
silicon	$-8.253\text{x}10^{-5}$	$3.694\text{x}10^{-7}$	-.14	no

Table 3

Tests of the hypothes that $b = 0$ for 9 elements.

References

1. Allen, C. W., "Astrophysical Quantities" Second edition. The Athlone Press, University of London. London 1963.
2. Anderson, T. W., "An Introduction to Multivariate Statistical Analysis" John Wiley & Sons, Inc., New York 1958.
3. Bond, B. B., "Spectrometric Oil Analysis." NARF-P-1 (Rev 6-67). NAVAIREWORK-FACPENS. Pensacola, Florida. 1967.
4. Graybill, F. A., "An Introduction to Linear Statistical Models, Volume I." McGraw-Hill Book Co., Inc. New York 1961.
5. Kittinger, Donald C. and John L. Ellis. ASD-TR-68-2, ASD, AFSC, Wright-Patterson Air Force Base, Ohio. 1968.
6. Marcus, Marvin and Henryk Minc. "A Survey of Matrix Theory and Matrix Inequalities." Allyn and Bacon, Inc., Boston. 1964.

DISTRIBUTION LIST

Lcdr Samuel Gordon Office of Naval Research (Code 463) Department of the Navy Washington, D. C. 20360	2
Commanding Officer Office of Naval Research Branch Office Box 39 FPO New York, New York 09510	1
Director Naval Research Laboratory Washington, D. C. 20390 ATTN: Technical Information Division	1
British Navy Staff (via NRL)	3
Canadian Joint Staff (via NRL)	2
B. B. Bond Materials Engineering Division NARF, NAS Pensacola, Florida 32508	2
Donald C. Kittinger Systems Engineer Group Research and Technology Division AFSC, Wright-Patterson AFB, Ohio	2
Dr. D. R. Barr Code 55Bn Naval Postgraduate School Monterey, California 93940	25
Dr. H. J. Larson Code 55La Naval Postgraduate School Monterey, California 93940	25
Defense Documentation Center Cameron Station Alexandria, Virginia 22314 ATTN: IRS	20
Library Naval Postgraduate School Monterey, California 93940	2
Dean C. E. Menneken Dean of Research Administration (Code 023) Naval Postgraduate School Monterey, California 93940	2

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

OBJECTIVE IDENTIFICATION PROCEDURES FOR THE NAVAL OIL ANALYSIS PROGRAM

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Research report

5. AUTHOR(S) *(First name, middle initial, last name)*

Donald R. Barr and Harold J. Larson

6. REPORT DATE September 1969	7a. TOTAL NO. OF PAGES 53	7b. NO. OF REFS 6
8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) NPS-55Bn55La9091A 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

A class of rules is developed for making decisions concerning whether a mechanical system may be failing, based upon spectroscopic analyses of the system's oil over a period of time. Some considerations that went into the development of these rules, including conclusions based upon studies of certain analysis records and experiments, are presented. It is indicated that these identification procedures should perform well in connection with a computerized analysis system, at least insofar as routinely monitoring the "well behaved" systems, while calling the attention of appropriate personnel to possibly discrepant systems.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Oil Analysis						
Decision Rule						
Prediction						
Multivariate Normal Distributions						
Spectrometric Oil Analysis						